UNCERTAINTIES IN COMPUTER SPECTROSCOPY FROM MACHINE LEARNING

NIKESH S. DATTANI, *Digital Technologies, National Research Council of Canada, Waterloo, ON, Canada.*

Computer spectrometers have recently been used to 'measure' the energy associated with the $C \rightarrow C^+$ ionization[a], the $Li_2(1^3\Sigma_u^+) \rightarrow 2Li$ separation[a], and the vibrational levels and vibrational spacings of $Li_2(1^3\Sigma_u^+)$[bc]. These were all cases involving up to 6 electrons. Up to at least 6 electrons, computer spectroscopy has been as accurate, almost as accurate, or in some cases *even more accurate*, than the best laboratory spectrometers, for measuring quantities such as ionization energies and dissociation energies, but we lag behind laboratory experimentalists when it comes to assigning uncertainties on our 'measurements': There is no easy way to do it accurately. Typically basis set incompleteness is the biggest source of uncertainty in computer spectroscopy. Energies are calculated with larger and larger basis sets, and then an extrapolation is done to approximate a CBS (complete basis set) energy. The uncertainty associated with this approximation can be estimated based on doing multiple extrapolations, with slightly different models, and looking at the spread of extrapolated values, but this is still a rather *ad hoc* way to estimate the uncertainty. Similar extrapolation schemes are also starting to become popular for estimating the correlation energy at a given basis set size.

My proposal is to develop a way to assign uncertainties to computer spectroscopy measurements in a way more similar to how it is done for laboratory instruments. Instruments often have a 'precision rating' based on the likelihood that the true value of the measured quantity is outside a specified window of precision around the number reported by the instrument. The measurement uncertainties reported in NIST's atomic spectra database are 1σ uncertainties, meaning that there is approximately a 1/3 chance of the true value being outside of the reported error bar. The precision rating can therefore be calibrated by doing many measurements, but for computer spectrometers we rely on building a big database and doing machine learning. I discuss the determination of uncertainties both for CBS extrapolations and for FCI energy estimations.

---

[a] Nike Dattani (2018) "0.06 cm^{-1} Discrepancy for $Li_2 \rightarrow 2Li$ and 0.994 cm^{-1} for $C \rightarrow C^+$ between Laboratory and Computer Spectrometers" Proceedings of the 73rd International Symposium on Molecular Spectroscopy.

[b] Nike Dattani (2017) "Computer Spectrometers". Proceedings of the 72nd International Symposium on Molecular Spectroscopy.

[c] Nike Dattani, Sandeep Sharma, Ali Alavi (2016) "Full CI Benchmark Potentials for the $6e^-$ System Li_2 with a CBS Extrapolation from aug-cc-pCV5Z and aug-cc-pCV6Z Basis Sets Using FCIQMC and DMRG". Proceedings of the 71st International Symposium on Molecular Spectroscopy.